Technological Advance in Cholesterol Medication Meets Physician Learning: A Non-Parametric Bounding Approach

Vincenzo Atella, Federico Belotti,
Jay Bhattacharya, Domenico Depalo*

May 29, 2018

Abstract

In this paper we investigate the relation between technological advance, in the form of an information shock, and changes in physicians and patient behavior. We focus on the case of statins, a medication used to manage high cholesterol levels, where newly produced evidence has altered the scientific consensus regarding side effect profiles. Early common wisdom held that statins may cause liver damage and therefore patients should be regularly tested for changes in liver enzyme levels. In 2010 GREACE, a large and influential randomized controlled trial, suggested instead that the medication could be continued despite elevated liver enzymes, and that patients need not be tested regularly for liver damage. We exploit this major informational technological shock to test how physicians prescription and testing behavior and patients adherence to therapy changed in response. We test our model using a unique dataset representative of the Italian population, that links patients to doctors over the period 2003-2014. We account for the possible non-random sorting of patients into treatment by exploiting an instrument which is assigned effectively at random. We employ a non-parametric bounding approach that takes into account the selection mechanism (unlike ordinary least squares) and permits conclusions about the entire population (unlike standard instrumental variables estimation). Our results show that doctors responded promptly to this technological shock.

JEL classification: I18, J18, C21
Keywords: Health status, Bounds, Technology

*We thank ... Replication files and additional results will be available at the webpage: http://sites.google.com/site/domdepalo/. The views expressed in this paper are those of the authors and do not imply any responsibility of their institutions. Corresponding address: Domenico Depalo, Banca d'Italia, Economics and Statistics Department, Via Nazionale, 91 - 00184 Roma, Tel.: 39-06-4792 5989, e-mail: domenico.depalo@bancaditalia.it

1 Introduction

Medical practice should be based on developments in medical science. Such developments may include new drug products, new knowledge about physiological mechanisms, and new information about existing technologies. Important new technological developments should in principle lead to changes in optimal medical practice by physicians. This is especially true if the new evidence convincingly contradicts prior common medical wisdom. Patients behavior such as adherence to prescribed therapy may also change in response to new technological developments. This may happen either because patients themselves become aware of the new evidence, or because changes in prescribed therapy induced by the technological development may improve clinical outcomes or make side effects less likely or less severe. In short, important new technologies and new medical information should disrupt medical practice. In this paper, we analyze the effects of an information shock that altered the scientific consensus on the side effect profile associated with statin medication, an important pharmaceutical product used to manage hypercholesterolemia.

Hypercholesterolemia is defined as high levels of cholesterol in the blood stream. High levels of low density lipoprotein (LDL) cholesterol is of medical concern because patients with high LDL levels are more likely to develop atherosclerosis and associated heart disease. Statin medications are among the most commonly used pharmaceutical products both because the prevalence of hypercholesterolemia is high, and because these medications have been shown in randomized trials to be safe and effective way to reduce serum LDL levels and reduce mortality. For instance, the 1994 Scandinavian Simvastatin Survival Study (among the earliest large randomized trials of statin medications) found that patients randomized to statins enjoyed a 35% reduction in LDL cholesterol and a 30% reduction in mortality hazard relative to placebo controlled patients

(Scandinavian Simvastatin Survival Study Group, 1994).

However, like every drug, statin medications can cause adverse side effects. In the case of statins, an increase in blood levels of liver enzymes like alanine aminotransferase (ALT) – which in some cases indicate liver damage – was historically seen as the most common side effect. An early randomized study found fifteen times higher rates of elevated liver enzymes among patients assigned to high dose statins relative to placebo or low dose statins (Bradford et al., 1991). Physicians viewed this evidence of side effects as salient; in studies of statin initiation, physicians were found to be reluctant to prescribe statin therapy because of the risk of hepatotoxicity (Rzouq et al., 2010). A finding of elevated liver function tests (LFTs) might lead a physician to recommend discontinuing statin therapy, lowering statin doses, or changing the statin molecule prescribed (Calderon et al., 2010). Given the concern about liver damage, until 2011 clinical guidelines suggested that physicians who prescribe statins for their patients should regularly test for changes in liver enzyme levels (McKenney et al., 2006; Gillett and Norrell, 2011).

In fact, even in the early days of statin medications, there were non-randomized studies that found that elevated liver function tests did not necessarily indicate permanent liver damage. A finding of elevated liver enzymes was often readily reversible (Mölgaard et al., 1991). Some observational evidence published in the mid-2000s suggested that this fear of liver damage was overstated.

An early 2004 study concluded, on the basis of a literature review, that routine liver function tests to help detect liver damage in patients on statins, was not necessary given that most cases of elevated liver enzymes were transitory (Sniderman, 2004). A series of observational studies of statin use among patients with pre-existing liver disease found the drug to be safe and effective, with no evidence of hepatotoxicity (Khorashadi et al., 2006; Segarra-Newnham et al.,

2007; Lewis et al., 2007; Bhardwaj and Chalasani, 2007). Evidence from around the world raised the idea that the link between statins, elevated liver enzyme levels, and permanent liver damage was not well established. For instance, a 2006 Italian study reviewed the evidence that statins cause liver damage, and concluded that though the authors found some evidence for a relationship, there was not enough to support a causal link (Conforti et al., 2006). An 2006 American expert panel of hepatologists concluded that statin use was largely safe in terms of liver side effects (Cohen et al., 2006). However, given the evidence from the early statin randomized controlled trials (RCTs), the general consensus in the medical community was that statins had the potential to cause this side effect (Bays, 2006; Clarke and Mills, 2006).

In late 2010, the GREACE study of statin medication use in a population of Greek patients with abnormal liver tests was published in *Lancet*, a leading medical journal (Athyros et al., 2010). The study showed that despite elevated liver enzyme levels, patients randomized to statin medications actually experienced *lower* rates of elevated liver enzymes relative to control patients, with a better cardiovascular response. This study was striking because its results were so unexpected given what most physicians believed about the effects of statins. If patients with liver disease show improvements in liver function after statin therapy is started, then the prior conventional wisdom that statins can cause low grade liver damage in patients without liver disease is most likely incorrect. This in turn means that a finding of elevated liver function tests in the context of statin therapy should not be taken as evidence that the statin medication is leading to liver damage. The GREACE results suggested that statin medications could be continued despite elevated liver enzymes, and hence that patients on statins need not to be tested regularly with LFTs (Pastori et al., 2015).[1]

[1]The GRAECE study did not contradict a previous finding that in some rare cases, statins can cause rhabdomyolosis, a serious condition involving muscle degradation that can lead to extensive liver and kidney damage. This rare side effect cannot typically be prevented or

Ultimately, this reasoning led experts to reverse the recommendations in previous guidelines. American practice guidelines were modified in 2014 to recommend that patients on statins did not need have their liver enzyme levels checked regularly. They suggest instead that liver function tests are needed only when statin therapy is initiated and when the statin dose or type of statin prescribed is changed (Bays, 2006). Though the Italian guideline for ordering liver function tests in the context of statin therapy has not changed, the GREACE study was widely known among Italian physicians.[2]

Our main aim with this study is to test how physician prescription and testing behavior changed in response to changing evidence about the side effect profiles of statin medication. This is important because, while there is much evidence of new technology adoption by physicians, there is little evidence about shocks in information about existing technology on physician behavior and patient outcomes. Given the history we recount above, we divide our data into three distinct periods: (1) an early period when most physicians believed that statins may cause liver damage, and hence regular liver function testing is necessary; (2) a period where some observational studies suggested that statins might not be the cause of liver function abnormalities in patients taking statins; and (3) after the publication of the GREACE trial, which established that statins may actually help patients with liver function abnormalities. A secondary aim of this paper is to understand how patient adherence to therapy changed in response to this changing evidence. Together, we aim to estimate how both doctors and patients react to new information in a discrete setting.

To do this, we analyze data from a large longitudinal sample of Italian patients who have been diagnosed with high cholesterol levels between 2004 and 2014. The dataset includes the identity of the primary care physician

detected early with regular liver function tests.

[2]Personal communication with Dr. Cortese, lipidologist at the University of Rome.

who manages the patient's statin prescription during this period. The data set includes biometric information, such as the measured LDL cholesterol level, as well as information on the identity and dose of statins prescribed at each time point, and information about the ordering of lab tests such as liver function tests.

To generate estimates that may be interpreted causally from this observational database, we employ recently developed econometric bounding methods that permit us to explore the effects of progressively stronger assumptions (Manski, 1990; Shaikh and Vytlacil, 2011). The bounding methods we employ – in contrast to more traditional point estimators employed by randomized and observational studies alike – permit us to measure the range of treatment effects that are consistent with observational data without the strong structural assumptions that are necessary to guarantee point identification.

The rest of the paper is organized as follows. Section 2 presents our data set. Section 3 illustrates a simple principal-agent theoretical model which helps the understanding the economic insights behind our empirical results. Section 4 provides a brief description of the recent econometric literature on bounding methods, and develops a small extension of these methods to continuous outcomes that we employ in this paper. Section 5 presents our quantitative results, and finally, in Section 6, we conclude providing some reflections on what our results suggest about the role of randomized evidence in medicine.

2 Data and Summary Statistics

In this section, we describe our dataset and inclusion and exclusion criteria, our measures of drug potency and patient adherence to therapy, and finally we present some important summary statistics about our population.

2.1 Italian Health Search Database

Our empirical analysis is based on data obtained from the Health Search Database (HSD), a longitudinal observational database collected by the Italian College of General Practitioners (SIMG) since 1998. The HSD contains patient level data from computer-based patient records reported by General Practitioners (GPs) throughout Italy. GP participation is on a voluntary basis, but nevertheless GP represented are in fact representative of the National Health Service (NHS) regional organization. The number of patients tracked in the HSD are in proportion to the size of the Italian adult population within each region (Fabiani et al., 2004).[3]

Patient data are linked through a unique anonymous identifier to drug prescriptions, clinical events and diagnoses, hospital admissions, and causes of death. Information available include prescription dispensing date and drug characteristics (Anatomical Therapeutic Chemical or ATC code which indicates the quantity and type of active ingredient and number of pills). Other observable characteristics are discussed in Section 2.3. We limit our sample to patient records collected between 2003 and 2014, inclusive, because prior to 2003 serum cholesterol level data were not accurately recorded by all GPs.

We selected patients based on two main inclusion criteria: *i)* patients who receive a diagnosis of "pure hyper-cholesterolemia" (or familiar hyper-cholesterolemia) and received at least one statin prescription between 2003 to 2012; and *ii)* patients born between 1925 and 1975, inclusive. This leads to a twelve year unbalanced panel, with 11 years of follow-up for the cohort of patients who were observed to be on statins in 2003 and only two years of follow-up for the 2012 cohort (that is, 2013 and 2014).

[3]Data have been collected routinely since 2000. More details on the representativeness of the database can be found at `https://www.healthsearch.it/` (Official website of the HSD project).

Since the initial sample consists of daily observations, after having constructed and transformed all variables used in the analysis, data have been collapsed to obtain patient-quarter level observations. Our choice of the patient-quarter as the unit of observation represents a compromise aimed at reducing the number of observations with a zero in drug consumption, while permitting relatively frequent changes in treatment and outcomes. The final sample is then organized as a quarterly unbalanced panel with data from 2004 to 2014, which consists of 112,331 patients (57,244 women and 46,493 men) for a total of 1,792,595 observations.

We do not exclude patients who initiated statin therapy prior to 2003 for two reasons. First, the half life of statin molecules in humans depends on the molecule, but in all cases we consider, is less than 20 hours (Plakogiannis and Cohen, 2007). Half-life, here, is a physiological property of a molecule, and indicates the amount of time it takes for half of the initial dose to be excreted from the body through the kidney, or metabolized into a biologically inactive form in the liver. Since the half-life of statin medications are so short, the effect of the drug on cholesterol levels (and other outcomes) for someone who initiates therapy before 2003 will depend primarily on the extent to which the patient takes the statin during 2003 (and not on drug-taking behavior before then). Second, including people who are already on therapy in 2003 may lead us to underestimate the effect of statin medications on cholesterol levels. Someone who has routinely taken statin medications for some time prior to the start of our sample in 2003 is likely to have a low measured cholesterol level at the start of the observation window. Continuing to take the statin medication through 2003 and after will not lead to a yet lower cholesterol level since the patient is already at his or her steady-state cholesterol level.

2.2 Measuring Drug Potency and Patient Adherence

We characterize the prescribed treatment regime according a univariate measure called the equipotency score. This score relies on the physiological fact that some molecules are more active than others in blocking cholesterol synthesis in the liver. This measure also accounts for the fact that higher doses of a molecule will have a more potent cholesterol synthesis blocking effect than lower doses. Table 1 shows this conversion formula for the three statin molecules that we analyze in this study (simvastatin, atorvastatin, and rosuvastatin). This table follows the work of Maron et al. (2000).

According to this table, a one unit increase in equipotency for the same molecule by doubling the dose. For example, going from 20 mg to 40mg of Simvastatin increases the expected LDL cholesterol reduction by one equipotency unit. Similarly, switching from 20mg of simvastatin to 20mg of atorvastatin or to 10mg of rosuvastatin leads to an increase in the equipotency score because rosuvastatin is more biologically active than atorvastatin, and atorvastatin is more biologically active than simvastatin. Analogously, switching from 40mg of simvastatin to 20mg of atorvastatin or 5mg of rosuvastatin maintains the same equipotency level.

A key characteristic of interest is a patient's adherence with the prescribed treatment regimen. This is important because a drug cannot have an effect if a patient decides not to take it. Given our data, there are various ways we could measure patient adherence.[4] In this study, we measure patient adherence to care using the Mean Possession Ratio (MPR) (Cramer et al., 2008; Atella et al.,

[4] According to the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), medication compliance is commonly measured by adherence and persistence to prescribed medical treatment. Adherence refers to the proportion of prescribed doses taken in the prescribed time interval, while persistence refers to the continued use of a prescribed therapy over time (Hughes et al., 2001).

2017). This is defined as:

$$MPR = \left(\frac{\text{Sum of days' supply for all fills in quarter}}{\text{Number of days in quarter}} \right) \tag{1}$$

The MPR is thus the fraction of days over the course of a quarter covered by all prescriptions for a drug. Low values of MPR are consistent with poor adherence because a patient who is poorly adherent to therapy does not need a refill prescription.

2.3 Other Observable characteristics

The Health Search Database also includes detailed information on patients clinical histories. As a summary of the general individual health condition, we construct the Charlson et al. (1987) index, a composite measure for the seriousness of diseases that increases as health conditions become worse (i.e., the assigned weight for AIDS is 6, the maximum, while for flu, it is 0, the minimum). We have also information about the presence of several chronic diseases, such as diabetes, hypertension, congestive heart failure, atrial fibrillation, vascular diseases, a history of cardiac bypass surgery or percutaneous coronary intervention (PCI), ischemic heart disease, or other cardiac conditions, which will be useful in order to control for comorbidities. In terms of diagnostic tests, we observe the number of patient specific prescribed tests by each physician for alkaline phosphatase (ALP), aspartate aminotransferase (AST), alanine aminotransferase (ALT), albumine and bilirubin. These tests represent our key variables for the empirical analysis as they allow to estimate the effects of informational shocks on patient prescription and physician adherence behaviours. Finally, a set of socio-demographic indicators includes age, sex, and the geographical region.

2.4 Summary Statistics

Table 2 reports the list of variables used in this analysis with their relative descriptive statistics by gender status. The dependent variable is the yearly rate of change in LDL-cholesterol at the patient level. As we can see in the table, men and women reduce their LDL cholesterol by an average of 2.4% and 2.6% (respectively) per year that they are in our sample. However, the median change is much lower, in the range of 1%. This marked difference between mean and median change suggests a left skewed distribution, denoting also a high heterogeneity in the effect of drug treatment on LDL cholesterol levels across patients.

In our selected population, 55.0% are female. Mean age is around 69 years for women and 65 for men. In accordance with the distribution of the Italian population, the majority of the patients are concentrated in the north of Italy (25.0% North-west, 20.0% North-east), with the center accounting for about 18.0%, and the south and the islands (Sardinia and Sicily) account for another 25%. The Initial level of cholesterol (that is, the first observed cholesterol level for a patient in our sample) is higher for women (129 mmol/L) than for men (118 mmol/L). The Charlson et al. (1987) index is similar by gender (about 1.0); in about one third of the observations also diabetes is present and hypertensions in two third of observations. Tests for aspartate aminotransferase and alanine aminotransferase are the mode among the available test and are prescribed about once per year (or 0.25 per quarter).

Table 3 reports the sum of all prescriptions recorded from 2004 to 2014 in our dataset by sex, dosage and active ingredient. This table thus reflects specific GP prescription patterns for simvastatin, atorvastatin, and rosuvastatin in Italy. As is evident in the table, the most prescribed active ingredients/dose combinations are Simvastatin 40mg, Atorvastatin 40mg and Rosuvastatin 20mg. Low dosage

statins are hardly prescribed, with some exception for Rosuvastatin 10mg which is actually at a high equipotency level.

3 Theoretical model

In this section, we propose a simple stylized model of physician beliefs about the efficacy and side effects of cholesterol medication, their optimal recommended choice of therapy given those beliefs, and finally patient adherence behavior in response to those recommendations. Our main aim is to model the effects of the information shock produced by the GREACE trial on physician beliefs to guide our empirical work in later sections of the paper.

3.1 Physician Objectives

We model physician beliefs about the effects of statin medication along two dimensions – the effect of statins on the cholesterol level (C) of a patient, and the likelihood of potential side effects (such as liver damage or muscle damage) produced by the medication. To simplify our theoretical modeling, physicians only consider the equipotency dose (Q) in their prescription recommendation, so the model elides the distinction between switching to a more potent molecule and switching to a higher dose of the same molecule. We also assume that side effects can be measured based with a single positive real valued number, SE, where larger values indicate either more or worse side effects for the patient (or both). For notational convenience, we leave implicit that all the beliefs and objective functions we discuss are conditioned on observable characteristics, X.

Higher serum cholesterol levels and more side effects are bad outcomes that physicians and patients want to minimize. To fit our discussion into the usual maximization convention, we introduce the following notation: $\overline{C} = -C$ and $\overline{SE} = -SE$. We will recast our discussion of physician objectives and beliefs

and patient objectives on $\overline{C}$ and $\overline{S}E$ rather than C and SE. Our focus in most of this discussion is on the decisions of doctors to recommend therapy, so we start with the physician utility function, with physicians serving as perfect agents for patients. In Section 3.6, we relax this assumption, and permit patient behavior to deviate from physician recommendations.

Physicians evaluate a combination of cholesterol and side effects outcomes with a von-Neumann-Morgenstern utility function $u(\overline{C}, \overline{SE}|Q)$, that is a nondecreasing function of both arguments. As a consequence, physician utility does not decrease when a patient's cholesterol or side effect levels decline:

$$\frac{\partial u}{\partial \overline{C}} \geq 0; \text{ and } \frac{\partial u}{\partial \overline{SE}} \geq 0. \tag{2}$$

3.2 Physician Beliefs

From a physician's (and patient's) point of view, prescribing (and taking) statin medications does not produce a certain outcome; rather, it produces a range of possible outcomes which are uncertain at the time of prescription.[5] This uncertainty is a key feature of decision making by doctors, and so decisions will be based on physician beliefs about the probability of cholesterol control and side effects, for a given equipotency dose Q: $F(c, se; Q) = P(\overline{C} < c, \overline{SE} < se|Q)$. These beliefs, F incorporate the scientific evidence regarding the clinical effects of statins as well as all other relevant information available to the doctor, such as patient clinical characteristics, X. In particular, F can be interpreted as incorporating physician beliefs about patient adherence to recommended therapy. We explore this issue further in Section 3.6.

Based on the randomized trial evidence we cite in Section 1, we assume that

[5]To keep the modeling effort simple and focused on the effect of physician beliefs about treatment efficacy, we do not explicitly model the principal-agent nature of the relationship between doctors and patients. In particular, we assume that doctors make recommendations that are in line with patient preferences. We relax this assumption in Section 3.6, and the paper more generally in that we permit patients to not adhere perfectly to prescribed therapy.

physicians believe that a higher equipotency dose will increase the probability of side effects and reduce the probability of a high cholesterol level (conditional on a given side effect level). We formalize this assumption as:

$$F(\overline{SE}|Q) < F(\overline{SE}|Q') \qquad \forall\, Q' > Q \tag{3}$$

$$F(\overline{C}|SE, Q) > F(\overline{C}|SE, Q') \quad \forall\, SE \text{ and } Q' > Q. \tag{4}$$

Equations 3 and 4 imply that, in expectation, increases in equipotency dose Q reduce cholesterol levels and increase side effects. Applying our notational convention, we have $\forall Q' > Q$:

$$E[\overline{C}|Q] < E[\overline{C}|Q'] \tag{5}$$

$$E[\overline{SE}|Q] > E[\overline{SE}|Q']. \tag{6}$$

In what follows, we will mostly rely on equations 5 and 6, which we view as mild restrictions, as they make a statement only about the first moment of the (C, SE) distribution, that are based on RCT evidence.

3.3 Changing Evidence and Changing Beliefs

We next model how the release of the GREACE trial changed physician beliefs. The GREACE trial evidence (primarily) revised the scientific view about the effect of statins on liver function damage, rather than the effect of statins on cholesterol. The latter effect was well established based on prior studies conducted in the 1990s and earlier. The belief shock we are interested in, then, is the change in the probability of a side effect, conditional on cholesterol level, for any given quantity Q. Applying Bayes' rule to the von-Neumann-Morgenstern

utility function, for any belief, F, we have:

$$F(\overline{C}, \overline{SE}|Q) = F(\overline{SE}|\overline{C}, Q)\, F(\overline{C}|Q). \tag{7}$$

Here, for a given dose Q, $F(\overline{C}|Q)$ is the marginal c.d.f. implied by F on serum cholesterol, and $F(\overline{SE}|\overline{C}, Q)$ is the conditional c.d.f. implied by F for the effect on side effects conditional on cholesterol level.

Let F^{POST} represent beliefs about the relationship between equipotency dose, cholesterol levels, and side effect profiles that a well informed physician will hold after the release of the GREACE trial. Correspondingly, let F^{PRE} represent the beliefs – based on the best available evidence in the literature at the time – about these relationships at the beginning of our analysis period (pre-GREACE).

We assume that the GREACE trial did not change beliefs about the effect of statins on serum cholesterol levels, or:

$$F^{POST}(\overline{C}|Q) = F^{PRE}(\overline{C}|Q). \tag{8}$$

The main findings of GREACE were focused on the hepatic side effects of statins, and in particular, that statins are less likely to cause hepatic side effects than was previously believed. Formally, we assume that

$$F^{POST}(\overline{SE}|\overline{C}, Q) < F^{PRE}(\overline{SE}|\overline{C}, Q),$$

which implies a first-order stochastic dominance relationship – F^{POST} dominate F^{PRE}. Combined with equation 8, this in turn implies that:

$$F^{POST}(\overline{C}, \overline{SE}|Q) < F^{PRE}(\overline{C}, \overline{SE}|Q). \tag{9}$$

For the purposes of the empirical application the main upshot of this analysis is as follows: a physician whose beliefs follow the literature before GREACE expects that a given dose of statins would produce more side effects than a physician whose beliefs are based on the post-GREACE consensus:

$$E^{PRE}[\overline{SE}|Q] < E^{POST}[\overline{SE}|Q]. \tag{10}$$

3.4 Physician Learning

So far, we have modeled the effect of a large one-time shock in beliefs, generated by the GREACE trial. In fact, physicians update their beliefs all the time on the basis of their experience treating patients. Here, we model this experience as a process of learning by doing, in which the more patients that a physician treats for hypercholesterolemia, the more experience the physician gets. The idea is that the more patients that a physician treats, the closer that physician beliefs will be to reality.

The natural question that arises in this context is whether the results of the GREACE trial reflect the true relationship between statin dose (Q), cholesterol levels, and side effects. Given the results of the GREACE trial, it should be clear that the pre-existing literature did not reflect reality, or else the results of the GREACE trial would have been quite different. In principle, though, the GREACE study is not necessarily definitive, and future studies could further refine scientific knowledge about these relationships.

For the purposes of this paper, we assume that GREACE has uncovered the true relationship, F^0, between Q, C, and SE:

$$F^{POST} = F^0. \tag{11}$$

We assume equation 11 because we believe – based on the current scientific

consensus cited in our literature review – that the GREACE trial represents a closer approximation to F^0 than did beliefs engendered by the evidence base prior to GREACE (F^{PRE}). Below, we discuss the consequences for our analysis if in fact it is found that the GREACE results are not correct.

We model the physicians learning process as a function of the number of patients treated. We adopt a Bayesian framework. Let N be the number of patients that a physician has treated for this condition and let $g(N)$ be a function with the following properties:[6]

$$\begin{aligned} g(0) &= 0 \\ \lim_{N\to\infty} g(N) &= 1 \\ g'(N) &> 0. \end{aligned}$$

We model physician beliefs, F^{doc}, as follows:

$$F^{doc}(\overline{SE}|\overline{C},Q) = g(N)F^0(\overline{SE}|\overline{C},Q) + (1-g(N))F^{PRE}(\overline{SE}|\overline{C},Q). \qquad (12)$$

As the number of patients increases, physician beliefs come closer and closer to the beliefs to the true effects of prescribing statins. Under assumption 11, this implies that, even before GREACE trial was published, physicians who treated a larger number of patients will behave as if statins do not produce liver side effects. Thus, for those physicians, the publication of the GREACE trial will have a smaller effect on their practice, with respect to physicians who treat fewer

[6]In principle, doctors update their beliefs based on the outcomes experienced by their patients that adhere to the therapy recommended by the doctor. If a doctor has many patients, most of whom ignore his advice, the doctor will not learn very much. One should think of N as the "effective" number of patients seen by the doctor. Doctors who are better at convincing their patients to adhere to therapy will then have a larger N, despite seeing fewer patients than doctors who do a poor job communicating the importance of adherence.

patients (and thus have beliefs closer to F^{PRE}), who will change their practice sharply after its publication. If physicians could observe the entire population, their belief would coincide with F^0 (by the property of $g(N)$). This is not possible, so we expect an effect for all the physicians. We test this implication of our model in the empirical section of this paper.

We recognize that equation 11 is a strong assumption that simplifies our theoretical analysis considerably. This assumption could be wrong in two distinct ways. First, it is possible that in truth, there are fewer side effects from statin use than were measured in the GREACE trial, so that:

$$F^{PRE} > F^{POST} > F^0.$$

In this case, the results of our theoretical analysis still hold with minimal modification, and the empirical results we report regarding the effect of practice size on physician practice are conservative (that is, they may be an underestimate of the true effect of the publication of the GREACE trial).

Alternatively, it is possible that future studies will conclude that the GREACE study was incorrectly optimistic about the side effect profile of statins. In that case, either

$$F^{PRE} > F^0 > F^{POST}$$

or even,

$$F^0 > F^{PRE} > F^{POST}.$$

In this second case, the interpretation of our result comparing doctors who treat many patients against doctors who treat fewer patients would need to be modified. If a future study finds that the GREACE was incorrect, and that statins do cause hepatic side effects, then we would be forced to conclude that physicians with large practices are currently behaving (and have been behaving

even before GREACE) on the basis of false beliefs about the side effect profiles of statins; they have learned nothing from their more extensive clinical practice. They are not updating their beliefs in response to the evidence in front of them.

3.5 Optimal Prescribing

We assume that doctors' utility follows the standard von-Neumann-Morgenstern assumptions, with the doctor taking the expectation given their own belief, F^{doc}, about the efficacy and side effects of statins. Doctors choose Q in a way that solves:

$$\max_{Q} \qquad U = E^{doc}[u(\overline{C}, \overline{SE})] \tag{13}$$

where the expectation is taken relative the doctor's beliefs.

The optimal Q^*_{doc} is chosen to maximize U with expectations set based on F^{doc}, which we assume is fixed near the optimum. The first order conditions are given by:

$$\frac{dU}{dQ} = \frac{\partial u}{\partial \overline{C}} \frac{\partial E^{doc}[\overline{C}]}{\partial Q} + \frac{\partial u}{\partial \overline{SE}} \frac{\partial E^{doc}[\overline{SE}]}{\partial Q} = 0. \tag{14}$$

The first order condition for this problem balances the marginal decrease in cholesterol level from a higher statin equipotency dose against the marginal increase in side effects.

While this procedure identifies the optimal dose given the doctor's expectations about the safety and efficacy of statin medication, if the doctor's beliefs do not match the truth ($F^{doc} \neq F^0$), then the observed cholesterol levels and side effect patterns will differ from what the doctor expects. In particular, the doctor expects to find side effect levels equal to $E^{doc}[SE; Q^*_{doc}]$ in response to prescribing the optimal dose, but in fact observes $E^0[SE; Q^*_{doc}]$, where the latter

expectation is taken with respect to F^0. Since $F^0 >_{fsd} F^{PRE}$ (see equations 9 and 11), we have:

$$E^0[SE|Q^*_{doc}] < E^{doc}[SE|Q^*_{doc}]. \tag{15}$$

In our empirical work, we indirectly test this implication of our model by analyzing the frequency at which doctors order a test for liver damage. The rationale beyond our approach is that as a doctor's beliefs come closer to F^0, we expect them to order fewer liver function tests since, as GREACE demonstrated, statins are unlikely to have such side effects.

As a doctor's beliefs tend toward F^{POST} – either because GREACE has been released or a doctor sees a large number of patients – it is easy to see that the optimal dose will increase, since $F^{POST} < F^{PRE}$. That is, our model predicts:

$$Q^*_{PRE} < Q^*_{POST}. \tag{16}$$

Since the optimal equipotency dose is predicted to be higher in the post-GREACE era (and for doctors whose beliefs are closer to F^0) we also predict lower cholesterol levels in the treated population relative to the pre-GREACE era.

This framework implies changes in physician prescribing on both the intensive and extensive margins, on an individual patient basis. That is, the GREACE shock both increased the optimal dose for patients already on a statin, and expanded the set of patients who optimally should be on statin therapy. One follow on implication from this observation is that patients selected for high does therapy in the pre-GREACE era were presumably at high risk of heart disease and other poor health outcomes caused by high serum cholesterol. This high

risk justified the risk of liver side effects that physicians thought, pre-GREACE, were a risk of statin therapy. The new information obtained about statins in the post-2006 era provides a nice opportunity to test this patient selection story.

3.6 Patient Adherence

Up to now, our modeling effort has emphasized the role that doctors play in prescribing therapy. Of course, patients play a crucial role in therapeutic decision making as well. If a patient decides against taking a medication, it does not matter how wise or well-informed the physician's prescription recommendation was – the therapy will have no effect on the patient's outcomes. So, to interpret any correlation between statin dose and the response of cholesterol level, it is crucial to consider how the recommended drug/dose combination, possible side effects, patient adherence, and the physician response to possible lack of adherence interact with one another.

Lack of adherence is particularly salient when considering treatment for a condition like hypercholesterolemia. High cholesterol is an asymptomatic condition, and outside of doctor visits, where serum cholesterol levels are measured, patients may not see any tangible, visible benefit from following the prescribed therapy. By contrast, though rare, some of the side effects that are caused by statin medication – such as muscle pain – may be painful. Patients may also incorrectly attribute other symptoms they are experiencing to presribed statin, even if the medical literature does not support such an attribution. In situations like this, adherence to prescribed statin medications is likely to be low.

Lack of adherence arises because patients have their own preferences and beliefs that physicians only imperfectly reflect. Though physicians in principle have more knowledge about the evidence regarding the effects of medications, patients certainly have their own ideas. If patients believed that their doctors

serve as perfect agents on their behalf, with full information about the preferences of the patients taken into account, there would be no problem of patient lack of adherence. In this sense, then, patient adherence to therapy depends on how much patients trust the recommendations of their doctor.

If doctors do a poor job in communicating the necessity of their recommendations to their patients, their patients will be less likely to adhere to the doctors' recommendations. Of course, some doctors are better than others in communicating with their patients. This observation motivates our use of physician-level adherence measures as an instrumental variable for adherence by particular patient (leaving out the patient's own adherence, of course).

Patient adherence is not a fixed constant, and in fact may change in response to changes in dose or prescribed medication. Our model predicts that physicians, whose goal it is to lower patient serum cholesterol levels to a safe level, in any given visit have two options to achieve this aim: switch the treatment to an *higher* level of equipotency or leave the prescribed regime at the same level of drug equivalence and place a great emphasis to the patient on the importance of adherence.[7] It is interesting to measure empirically, outside of randomized settings (where adherence is always strongly encouraged and often tightly enforced), the relationship between equipotency dose and serum cholesterol.

When patient adherence is taken into account, our theory predicts that changes in recommended equipotency level have a non-monotonic effect on serum cholesterol levels. When a higher dose is recommended by a doctor who is poor at communicating the importance of adherence, the patient might respond by consuming less than the recommended amount, and perhaps even less

[7] Given the array of stain options available to physicians, it is possible for physicians to recommend a lower prescribed dose in milligrams, while simultaneously recommending a switch to a more potent molecule. In net, this will increase may increase equipotency class, even though the dose in milligrams decreases. To simplify the analysis, we focus on changes in equipotency dose and do not separately analyze the effect of changes in milligram dose or recommended molecule.

than the previous effective dose. In that case, cholesterol might rise. A higher recommended recommended dose by a doctor who is good at eliciting adherence will, according to our theory, reduce the patient's cholesterol level. Information shocks like GREACE (which changed the expectations that doctors have about side effects) can intensify the incentives that physicians have to increase the dose or push to increased adherence. We examine these issues directly in the empirical section of the paper.

4 Methods

In this section we review the methodologies that allow us to test the economic model of Section 3. We first define our main parameter of interest – a population average treatment effect – then discuss the various empirical strategies we use to identify it.

4.1 Population Average Treatment Effect

We employ the standard model of potential outcomes, in which the outcome Y (such as reduction of cholesterol) is observed either under treatment (Y_1) and non-treatment (Y_0), but not both (Roy, 1951). In this notation, the subscript on Y indicates treatment status ($d \in D$ takes value 1 for treatment and 0 for non-treatment). The population average treatment effect (ATE) is given by:

$$\begin{aligned} \Delta &= E[Y_1] - E[Y_0] \\ &= E[Y_1|D=1]\,P[D=1] + E[Y_1|D=0]\,P[D=0] \\ &- E[Y_0|D=1]\,P[D=1] + E[Y_0|D=0]\,P[D=0], \end{aligned} \tag{17}$$

If we could observe the outcome under the two states of the world for a randomly chosen individual (Heckman et al., 2006), the identification of ATE would be

straightforward. In fact, the sampling process is informative only about the outcome under treatment for treated individuals ($E[Y_1|D=1]$) and about the outcome under non-treatment for non-treated individuals ($E[Y_0|D=0]$). The main challenge of the analysis is thus recovering the outcome for treated patients had they not been treated, or $E[Y_1|D=0]$, and vice versa for non-treated patients.

We represent the average treatment effect in terms of the following index function model (Vytlacil, 2002):

$$\begin{aligned} Y_d &= r(D) + \epsilon_d \\ D^* &= s(Z) - v \qquad D = 1(D^* > 0), \end{aligned} \tag{18}$$

where D^* is the selection rule that determines who is treated, on the basis of a binary variable Z (our instrument) which determines treatment, but not the outcome; ϵ_d and v are unobservable disturbances. An example of Z is the communication skill of the doctor. By contrast, the doctor's clinical ability is not a valid instrument since it presumably has a direct effect on the patient outcomes, Y.

4.2 Empirical strategies

We divide our discussion of our empirical methods into three sections that vary based on which assumptions we are willing to impose on the relation between ϵ_d and v. In the first, we discuss the methods to recover the *p*opulation average treatment effect under assumptions of exogeneity (ordinary least squares). In the second section, we discuss methods to recover the *l*ocal average treatment effect (LATE) if treatment assignment is endogenous (instrumental variables methods). In the third, we discuss methods to recover the *p*opulation average treatment effect under endogenous treatment assignment (bounding methods

with an instrument).

Recovering the ATE under exogeneity

In the simplest case, when $\epsilon_d \perp v$ we have that $E[Y_d|D=1] = E[Y_d|D=0]$. It then follows that $E[Y_1|D=0] = E[Y_1|D=1]$ and $E[Y_0|D=0] = E[Y_0|D=1]$, which can be substituted in equation 17 to derive $\Delta = E[Y_1|D=1] - E[Y_0|D=0]$. This parameter is valid for the entire population, but it is biased if treatment assignment is endogenous (that is, doctors assign treatment based on information they observe but that we do not).As such, it is incoherent with our stylized economic model. We present these results as baseline model to understand the effect of measuring the LATE rather than the ATE, and of accounting for endogenous treatment assignment.

Recovering the LATE under endogeneity

A major theme of Section 3 is that the sample of treated individuals is not completely at random, in the terminology of Little (1995). Physicians recommend therapy on the basis of variables that both we and they observe ("observables"), as well as on variables that they observe, but we do not ("unobservables"). In the remainder of this section, we take the position that conditioning on observables is not sufficient to solve the selection problem.

Suppose we have an instrument, Z, in the sense described in equation 18. We assume (without loss of generality) that Z and D are positively correlated. An instrumental variables estimator that is robust to sorting into treatment meets the following conditions (Imbens and Angrist, 1994; Angrist et al., 1996):

1. potential outcomes are unrelated to the treatment status of other workers (also known as stable unit treatment value assumption, or SUTVA);

2. the instrument is correlated with the treatment indicator, D;

3. the instrument does not affect the outcome (that is, Z does not enter the the Y_d equation in equation 18), and

4. monotonicity, so that $D_1 \geq D_0$ for each patient, where D_z is the potential treatment with the instrument $Z = z$. Under this assumption, there are exactly three groups: "always takers" who always treated no matter the value of Z, "never takers" who are never treated no matter the value of Z, and "compliers" who are treated if $Z = 1$ and are not treated if $Z = 0$.

Under these hypotheses, we identify

$$LATE_c \equiv \frac{E[Y_1|Z=1] - E[Y_0|Z=0]}{E[D_1|Z=1] - E[D_0|Z=0]} = E[Y_1 - Y_0|D_1 > D_0]. \tag{19}$$

While the estimate of the treatment effect is robust to endogeneity, it does not recover the population ATE. In fact, the IV estimator identifies a *l*ocal average treatment effect ($LATE_c$), valid for the subpopulation of patients who are induced to be treated by a change in the instrument. Patients belonging to this "complier" subpopulation cannot be indentified on the basis of the observable covariates, although the distribution of their characteristics can be calculated (Imbens and Rubin, 1997). Thus, when analysts are interested in a treatment effect *for the entire population*, as we are, IV-LATE is not be the best option.[8]

Recovering the ATE under endogeneity

Our goal in this section is to describe the empirical methods we adopt to measure the population ATE when treatment assignment is endogenous. Identifying the population average treatment effect is challenging because Y_0 is never observed

[8]IV-LATE would identify the population ATE if the treatment effect of statins can be assumed homogeneous within the whole population, or when the entire population consists of compliers, or patients do not self-select on the basis of their idiosyncratic returns (Heckman, 1997; Blundell et al., 2005), or some external information excludes specific subpopulations (Oreopoulos, 2006). None of these cases can be safely assumed in our context.

for always-takers ('at') and Y_1 is never observed for never-takers ('nt'). Consequently, we cannot identify, given reasonable assumptions, the local treatment effects, $LATE_{at}$ and $LATE_{nt}$.

The following decomposition is useful (Chen et al., 2017):

$$ATE = \pi_{at}\, LATE_{at} + \pi_c\, LATE_c + \pi_{nt}\, LATE_{nt}, \tag{20}$$

where π is the share of the attached component.

In this paper, we review some useful assumptions that allow us to recover the population ATE. These assumptions impose a cost: we cannot point-identify the ATE. Instead, we identify a set of possible ATEs that are consistent with our data, in the form of 'bounds' on the population ATE(Manski, 1990).

To fix ideas, consider the simplest possible example. Suppose we know, based on outside information, that Y is contained in a fixed set: $Y \in [k_0, k_1]$ with $k_0 \leq k_1$. Of course, we observe only one of the two potential outcomes, Y_1 or Y_0, for each patient. In this case, we know that even in the unobserved state of the world, the outcome would have been *at least* equal to k_0, thus identifying a lower bound on the outcome, and *at most* equal to k_1, thus identifying a upper bound on the outcome. Therefore, the upper bound on the treatment effect is equal to the difference between the upper bound of the outcome under treatment and the lower bound under non-treatment (Manski, 1990). Upon substitution into equation 17, it follows that

$$Lower \quad : \quad E[Y_1|D=1]\,P[D=1] + k_1\,P[D=0] - \{k_0\,P[D=1] + E[Y_0|D=0]\,P[D=0] \tag{21}$$

$$Upper \quad : \quad E[Y_1|D=1]\,P[D=1] + k_0\,P[D=0] - \{k_1\,P[D=1] + E[Y_0|D=0]\,P[D=0] \tag{22}$$

Although these bounds are very useful to understand how set identification works, they are usually very large and always cross the midpoint between k_1

and k_0. The reason this bound on the ATE so wide is that we assume so little. As a general rule, the stronger the assumptions one is willing to make, the smaller the width of the bounds.

In this paper, we consider imposing stronger assumptions to narrow the identified set along the lines of Shaikh and Vytlacil (2011); Bhattacharya et al. (2008, 2012); Chen et al. (2017). The first three papers consider a binary outcome variable and impose an individual level monotonicity condition similar to that of the IV-LATE. Chen et al. (2017) consider continuous outcomes and relax the monotonicity assumption.

More precisely, besides the IV-LATE conditions of Imbens and Angrist (1994); Angrist et al. (1996) that we enumerate above, we impose:

1. $Y \in [K_0, K_1]$;

2. $E[Y_1|S] \leq E[Y_0|S]$ for $S \in at, nt, c$ (or the opposite);

3. $E[Y_0|at] \geq E[Y|Z = 0, D = 0]$ and $E[Y_1|nt] \leq E[Y|Z = 1, D = 1]$, or the reduction of cholesterol under non-treatment for always-takers is smaller than for patients for whom $Z = 0$ and $D = 0$ (which consist of compliers and never-takers). Analogous reasoning applies to the second inequality.

The last condition deserves discussion. We impose it for two reasons. First, this condition is consistent with our economic model. Always-takers have no clinical alternative to statins, or else their cholesterol would be higher than the rest of the population. So they benefit on net from taking statins, and will do so regardless of the value of the instrument. Never-takers, by contrast, can substitute healthier behaviors (e.g., exercise) for drugs, so the balance between benefits and costs is more complicated (Atella et al., 2017). If all this is correct, the pre-treatment conditions of always takers (never takers) should be worse (better) than for other sub-populations. In the empirical application (Section 5),

we successfully check all these conditions. Second, the monotonicity assumption is directly testable, and in the appendix, we show that our data meet the test (Chen et al., 2017).

The bounds obtained imposing these assumptions are:[9]

$$
\begin{aligned}
&\text{if:} && E[Y|Z=1] - E[Y|Z=0] > 0: \\
LB &= && E[Y|Z=1] - E[Y|Z=0] \\
UB &= && E[Y|D=1,Z=1] - E[Y|D=0,Z=0]
\end{aligned}
$$

$$
\begin{aligned}
&\text{if:} && E[Y|Z=1] - E[Y|Z=0] < 0: \\
LB &= && E[Y|D=1,Z=1]\,E[D=1|Z=1] - E[Y|D=0,Z=0]\,E[D=0|Z=0] + \\
& && + K_0\,E[D=0|Z=1] - K_1\,E[D=1|Z=0] \\
UB &= && E[Y|D=1,Z=1]\,E[D=1|Z=1] - E[Y|D=0,Z=0]\,E[D=0|Z=0] + \\
& && + \min\{E[Y|D=0,Z=1], E[Y|D=1,Z=1]\}\,E[D=0|Z=1] + \\
& && - \max\{E[Y|D=1,Z=0], E[Y|D=0,Z=0]\}\,E[D=1|Z=0] \qquad (23)
\end{aligned}
$$

In this paper, inference is conducted using the method by Kreider and Pepper (2007), that is based on nonparametric bootstrap. We discuss these issues further in the appendix.

5 Results

In this section, we describe the results of our empirical analyses aimed at exploring the effects of changes in recommended dose and of the technological shocks in the medical literature about the side-effect profile of statin medication.

[9] Bhattacharya et al. (2008, 2012) discuss the relation between these bounds and those obtained by Manski and Pepper (2000) imposing Monotone Treatment Response (MTR). Depalo (2017) shows the one-to-one relation between the heterogeneity in the treatment effect in the population and width of bounds.

5.1 Data Subsetting and the Three Eras in Cholesterol Management

Our theoretical model (Section 3) makes several predictions that we test now. In most of the section we are interested in $\Delta = Y_1 - Y_0$. As a key idea underlying the economic background is that adherence to therapy exerts a positive impact on the reduction of cholesterol, we begin with this prediction. We move on by testing the relation between an increase in equipotency dose (Q) and cholesterol: this is important because before GREACE it was believed that increasing the equipotency dose reduced the cholesterol level but increased the side effects. Our model predicts that physicians can learn thanks to new information and/or the opportunity to visit more patients. If the GREACE trial altered the scientific consensus on this subject, then the clinical characteristics of the treated population should have changed over the treatment eras, as should have been the relationship between testing and switching behaviour. Even though this is true for all physicians, those visiting several patients should have acted more coherently with the information that was later provided by GREACE.

To test these predictions, we divide our analysis into three time periods (before 2006, 2006-2010 inclusive, and 2011+) based on information in the literature on the effect of cholesterol testing on liver function. Recall that in the early period, the conventional wisdom was that statins may cause liver damage in some cases. In the middle period, there was some non-randomized evidence that the effect of statins on liver function might not be directly causal. Finally, in the late period (after the GREACE randomized trial) statins were known to protect against liver dysfunction. Therefore, all our analyses are conducted separately on those three sub-periods.

A key point here is that the nature of the technological advance involved new information about the properties of an existing drug, rather than any physical

change in the molecule itself or its formulation. Thus, there is no reason to expect a change in the effect of drug on cholesterol reduction except through changes in physician and patient patterns in the use of drug.

In our point-identified results, our conditioning set includes the Charlson index (a composite comorbidity score explained in Section 2), age, sex, geographical region, and indicators of the presence of chronic diseases like diabetes, hypertension, congestive heart failure, atrial fibrillation, vascular diseases, a history of cardiac bypass surgery or percutaneous coronary intervention (PCI), ischemic heart disease, or other cardiac conditions. We include this extensive set of covariates in these regressions to reduce the chance of bias due to observed differences between patients assigned different statin molecule/dose combinations.

For the bounding estimators, by contrast, we do not condition on observable covariates, and instead verify that the assumptions imposed to generate bounds are met. (Please see Section A). Pepper (2000) presents an illuminating discussion about the different role of covariates in point and set identifications. For traditional point-identified estimators, namely OLS and IV, covariates generally serve the purpose of reducing bias due to relevant observable differences between the treated and untreated groups. In the bounding estimators, covariates are used to define subgroups of the population for whom separate treatment effect estimates are of some clinical interest; there is no bias problem as long as the assumptions underlying the estimator are met.[10]

[10]In an online appendix, we calculate our bounding results separately for men and women because it is in principle possible that the physiological effect of statins on cholesterol, as well as medication adherence, differ by sex. In practice, consistently with results from RCTs, we find little difference between the sexes in our results, so we present combined results here to save space.

5.2 Physician Habits as an Instrumental Variable

While it may be tempting to compare outcomes for patients on different treatment status, this procedure is not correct, because the set of treated and untreated patients is not randomly assigned. In fact, the status depends on a variety of clinical and economic factors known to both patients and physicians, but not always known to researchers.

To address this problem, we propose an instrumental variable that is motivated by our theoretical model – the physician-specific share of patients who are treated (please see Section 5.3). For a given treatment, this instrument is calculated for each patient by calculating the behavior of all the other patients that the patient's physician manages. This calculation purges any mechanically induced correlation between the instrument and the outcome thus being exogenous, and is consistent with the stable-unit of treatment value assumption (SUTVA) condition of Imbens and Angrist (1994).[11] In all our exercises, we empirically find that our instrument is strongly correlated with a patient's own behavior (the F-Stat from first-stage suggested in Bound et al., 1995 much greater than 10).

One possible concern about the instrument is that the selection process that matches patients to GPs is not random. However, in the Italian medical care system, the choice of general physician is strongly influenced by the location of the physician and the patient. General practitioners are not allowed to reject a patient for service on the basis of the severity of the patient's medical condition. Different values of our instrument, then, are more likely to reflect distinct practice styles of the practitioners (due perhaps to differences in their communication skills), rather than unobserved differences in physician behavior due to the underlying condition of the patient.

[11] The SUTVA implies that potential outcomes for each person i are unrelated to the treatment status of other individuals (Angrist et al., 1996).

We use a binary version of our instrumental variable, consistent with the description of the estimator we adopt in Section 4. To generate this binary instrument, we calculate the average value of the continuous version of our instrument. Then, we assign our binary instrument a value of one if the patient sees doctor who has an average switching rate above the mean switching rate for all doctors in the sample, and zero otherwise. This approach may be understood as a comparison of each physisian's practice style with relative to the community norm.

5.3 Treatment Adherence Results

A basic assumption of our model is that the adherence to therapy has a positive impact on the reduction of cholesterol: a patient not taking the prescribed statin will see little change in their cholesterol level. In this section, we conduct a test of this assumption.

In Table 4 we report the bounds of Section 4 (using the correction for the finite sample bias proposed by Kreider and Pepper, 2007), as well as OLS and IV estimates of the treatment effect. We also include some useful descriptive statistics: (log) cholesterol level change after six months (Y) conditional on adherence during this period of more than 75% ($D = 1$) and adherence during this period of less than 75%. We report these results for each type of statin molecule (for brevity we discuss only the agregate results from the row 'All'), as well as for each of the three treatment eras. The numbers in this table are all negative because these drugs tend to reduce cholesterol levels.

The OLS estimates tend to show small changes in LDL cholesterol levels correlated with patient adherence. Adherent patients experience (roughly) a 2% reduction in cholesterol levels relative to non-adherent patients. This is entirely consistent, not surprisingly, with the mean changes observed in the

right-most columns of the table. Though potentially biased because of the endogeneity of adherence behavior, the OLS results are interesting because they represent an estimate of the population level average treatment effect. The OLS treatment effect is smaller in later treatment eras (3.5% before 2006, about 2.5% between 2006 and 2011 and 1.5% after 2011): this is consistent with the idea that the set of people under statin treatment expanded as the fear of liver damage diminished, namely those with worse unobservable health conditions.

The IV results uses the instrument described in Section 5.2. As is the case with the OLS results, we find that adherent patients see a greater decrease in their cholesterol than non-adherent patients. The reduction from IV is never smaller than that from OLS, thus supporting the idea that the latter is biased from self-selection.

The IV results are interesting because they address the endogeneity of adherence, but the drawback is that it is consistent only for the local average treatment effect, rather than the population average treatment effect. In this case, the local average treatment effect applies to the set of patients who adhere to therapy because they are seeing a doctor whose other patients comply at a rate higher than the overall adherence rate in the population, but would not adhere otherwise. There is no reason to think that the treatment effect of adherence in this "complier" population should coincide with the population treatment effect. This population is not particularly interesting to clinicians since it is impossible in general to identify whether any particular patient belongs to it *a priori* (or indeed *ex post*). Because it is not clear who these compliers are, we should not interpret the changes in the IV estimates in earlier vs. later treatment eras as the response due to a technological shift for the population of patients suffering from hypercholesterolemia, nor we can say who was affected by this shift.

Finally, the BSV bounds in Table 4 are fairly wide. The lower bounds show

reductions in cholesterol on the order of 50%: this reduction may seem large, but in fact are consistent with the effect size measured in the clinical trials (Scandinavian Simvastatin Survival Study Group, 1994). These bounds shift from more to less effectiveness (from a more to a less negative parameter of treatment effect). Although the differences over time are always negligible –which is coherent with the fact that we purged from the unobservable components–, in relative terms they are more sizable at the upper bounds than at the lower bounds. As going from the first to the second era, the upper bound shrinks from 3.8 to 2.1% (a difference of about 2 percentage points); from the second to the thrid era the reduction is much smaller (from 2.1 to 1.6%, a difference of about 0.5 percentage points). Since these bounds are robust to non-random sorting (unlike OLS) *and* apply to the entire population (unlike IV), these results suggest a relative more important heterogeneity on the upper bound of the treatment effect (i.e., where the reduction of cholesterol is smaller).

Overall, the results in Table 4 are consistent with our hypothesis that adherence to therapy leads to lower cholesterol levels in the population at large.

5.4 Effect of Switching Equipotency Levels on Cholesterol

In this section, we explore the effect of changes in equipotency dose on reduction of serum LDL cholesterol. In one sense, one might think that the answer to this question is available from the numerous randomized trial data, which show that increasing statin dose leads to substantial reductions in serum cholesterol levels reported in Table 1. However, the story in actual practice is complicated by two phenomena that are not present in the randomized trials (Manski, 2013). First, the set of people who are treated in the population include many patients who have health conditions that were not represented in the initial trials (Deaton and Cartwright, 2017). Though subsequent RCTs may be conducted that expand

the set of patients exposed to treatment (such as the GREACE trial which included patients with liver disease), physicians are free to prescribe statins to patient groups not represented in the trial (and often do). Thus, the treatment effect in the population may differ from what is observed in the trials.

Second, the set of people who participate in the randomized trials are typically much more likely to adhere to prescribed therapy than are patients in the population at large (Frolkis et al., 2002). In trials, investigators regularly remind patients to stick to the therapy they have been randomly assigned; low adherence presents problems for investigators in calculating the treatment effect of the drug, so they take actions to increase adherence. This is less common in real world treatment settings, where patients make their own decisions about whether to follow their doctors' recommendations.

Hence, it is important to discuss the consequences of physicians changing (namely, raising) the prescribed equipotency level of the statin medication. These results are presented in Table 5. A doctor might decide to recommend a higher dose, for instance, if the current dose has not produced a sufficiently large reduction in LDL cholesterol, *and* there is no evidence of side effects or lack of adherence at the current dose. In such a case the switch will produce a decrease in LDL cholesterol. This is in fact what we empirically observe in all three treatment eras, and with all three estimators. Let us consider the estimators robust to unobserved components. When physicians raise the prescribed effective dose, then, both compliers and the population at large enjoy decreases in serum LDL.[12] Over time, the results from Table 5 are remarkably

[12]Even if the reader is skeptical about our choice of instrument, we argue that a violation of our IV assumption would likely tend to produce an attenuation bias on our results in this psecific analysis. Suppose that the physician's propensity to increase the dose is in fact correlated with a patient's unobserved health status. In the case of statins, one of the few reasons why a physician might recommend switching to an alternate therapy is that the current molecule and dose is not producing a sufficient reduction in LDL cholesterol level. Thus switches will be most likely to occur (and physician propensity to switch will be highest) when the observed change in cholesterol level is smallest.

stable, overall for set identification. However, to gain further insight on these results, notice that the IV is slightly increasing over time, thus the subpopulation of compliers may include patients with worse characteristics in later eras than in earlier eras. Since the medication is prescribed by physicians, having an in-depth look into the subpopulation of compliers is a first important test on how practitioner changed their behaviour, according to the new evidence. We do this in the next section. For completeness, the OLS estimates an ATE that is systematically biased towards smaller effectiveness of the treatment, like for the case of adherence.

These results are indistinguishable from those in Table 4. If anything, in relative terms the lower bounds are identical, whereas the differences on the upper bounds are slightly increasing over time. This could be due to the RCTs leading to more patients adherent (a claim that we successfully check in the next subsection). This is coherent with our claim that there is no reason to expect a change in the effect of drug on cholesterol reduction except through changes in physician and patient patterns in the use of drug. Since the decision about whether the patient should switch to a different dose is made by the physicians, in what follows we take their viewpoint.

The results in Table 5 do not mean that increasing the effective dose would always be effective, and that therefore all physicians should recommend that their patients do so. Rather, the results imply that physicians are good at predicting when an increase in the prescribed effective dose would not lead a patient to adhere less to therapy, and thereby gain the full benefit from a recommendation of a the higher dose.

5.5 Comparing Always Takers, Never Takers, and Compliers Over Treatment Eras

In this section we discuss how the clinical characteristics of the treated population has changed over the treatment eras. Our expectation is that increasingly more patients with worse characteristics are treated in later eras than in earlier ones.

Table 6 presents a wide variety of health and health economic variables for the patients in the sample. In addition to conditioning on the era of treatment, we divide the sample up into four groups based on their values of the instrument (Z) and whether doctors recommended a change in the equipotency level of the prescribed medication (D). These four groups are the "always-takers" (who has a change in equipotency recommended despite being seen by a doctor who recommends equipotency changes less frequently than average; we use the terminology in Angrist et al., 1996), "never-takers" (who do not have a switch recommended, despite being seen by a doctor who recommends equipotency changes more frequently than average), as well as "compliers". The last group consists of two distinct types of patients: (1) patients who are seen by doctor who recommends equipotency changes more frequently than average and who have a switch recommended; and (2) patients who are seen by doctor who recommends equipotency changes less frequently than average and do not have a switch in their recommended prescribed regimen. The first of these two "complier" groups include "always-takers" mixed in with the actual compliers, while the second of these two include "never-takers" mixed in. Thus the clinical characteristics we report in this table for these two "complier" groups are weighted combinations of actual compliers and some other population (Imbens and Rubin, 1997).

Two dimensions are relevant for the comparison: we can fix the period and

compare different sub-population or we can fix the sub-population and compare different periods. With respect to the former, on many important clinical and demographic characteristics, the "never-taker", "always-taker", and "complier" groups are indistinguishable from one another on average (Table 6). The characteristics where we observe clinically small differences include: age, sex, the number of liver function tests ordered (serum albumin, ALT, AST, blirubin), anemia prevalence, asthma prevalence, body weight or obesity prevalence, exercise habits, chronic obstructive pulmonary disease prevalence, HDL and LDL cholesterol levels, history of cardiac bypass surgery or balloon angioplasty, history of cancer, diabetes prevalence, atrial fibrillation prevalence, hypertension prevalence, ischemic heart disease prevalence, heart attack history, and history of vascular disease. This lack of differences across these groups can be summarized nicely by the fact that the distribution of Charlson et al. (1987) Index scores (which is a summary measure of the presence of chronic diseases in a patient) does not differ across the four groups in a clinically signficant way. The "always-takers" do have higher levels of adherence, higher expenditures on drugs, and higher total medical expenditures, than the "never-takers", but this just indicates that they tend to comply more with doctor orders.

One important consequence of this (lack of) finding is that trying to develop a model to predict whether a doctor is likely to recommend a swtich in therapy conditional on these observed clinical and demographic characteristics (perhaps using machine learning methods) would be a difficult challenge. There is clearly too much clinical overlap between the groups to permit an easily observable distinction. Doctors recommend switches in equipotency dose on the basis of characteristics that we do not observe, despite the fact that we observe so many clinical characteristics.

Also, from Table 6 it is clear that the adherence increased for all the types of

patients, thus confirming our reading that after the new information provided by GREACE the adherence rate increased.

By contrast with the results across patient types, Table 6 shows substantial change in the clinical characteristics of the population in the direction of sicker patients under treatment over time. For all subgroups of patients, health conditions (either in general, such as the Charleson Index, or those specifically related to LDL cholesterol) become worse, while the number of liver function tests per quarter decreases. This trend is particularly striking in the post-GREACE trial era. Coherent with the first main prediction of our model, as new information about the side effect profiles of statin became available, physicians update their beliefs so that the set of patients under treatment expanded to include patients who were previously thought to be poor candidates for the drug.

5.6 Relationship Between Testing and Switching

We continue to investigate the learning mechanism of physicians by looking at the probability of a change in recommended equipotency level as a function of number of liver tests (Table 7). For all patients (see the 'Overall' column), the probability of a change in therapy is an increasing function of the number of tests: patients who take at least four tests are more likely to switch than patients who have their liver function tested no more than three times. This is consistent with both the pre-GREACE guidelines and the post-GREACE American guidelines, which suggest testing liver function tests at the time of a recommended switch in therapy. Our results suggest that Italian doctors are following these guidelines, at least on average.

If we fix the period of analysis and compare different groups of switchers, patients who are prescribed a higher equipotency dose undergo a larger number of liver function tests. This is consistent with physician concern that raising

the dose, or prescribing a stronger molecule may result in a higher risk of liver side effects. Higher equipotency levels may be more effective, but physicians are concerned that they may also be more damaging.

Next, we examine changes across treatment eras, holding fixed switcher groups. This "vertical" comparison is motivated by the information shock provided by GREACE, as well as the observational evidence piling up before 2010.

For all groups of patients, there is an abrupt change in "testing behaviour" that is evident in the post-GREACE era. For example, with respect to the baseline category of no liver function tests, the probability of switching the therapy to a lower class of equivalence with four tests was larger than 2% (0.02138) before 2006, compared with less than two percent between 2006 and 2010 (0.01839), or a decrease of about 15%;[13] after the results of the GREACE trial became available, the probability of switching conditional on at leat four tests went further down (again, with respect to the baseline category) to about 1%, or a decrease of 30%. Similar results can be seen for patients switching to an higher equipotency class. The probability of switching the therapy to an higher equipotency class with four or more liver tests was slightly smaller than 15% before 2006, and decreased marginally in the following period (a decrease equal to 3.5%), but more strongly after GREACE (a drop as large as 25%). Clearly, the GREACE information shock affected liver function testing behavior for all groups of patients, with those switching to a lower equipotency class more affected. This pattern of results is consistent with the idea that Italian GPs changed their practice in response to new scientific information, and relatively quickly.

By and large we could not reject the first main prediction of our model that after a change in clinical evidence, there is a change in beliefs of the doctors.

[13] Given the large difference in the absolute values of the probability, we find the comparison of these relative risk more fair than the comparison of the attributable risk. For a more thorough discussion on this kind of comparisons and for similar conclusions, see Manski (2009).

5.7 Physician Experience and Physician Learning

The last prediction from Section 3 is that the larger the number of patients cured by a given doctor, the larger the opportunity for the physicians to update their beliefs such that they become closer to the truth. If so, physicians are less worried by the liver test of their patients because *thanks to their learning* they understand that the abnormal values are not the consequence of the statins (from Section 3 recall that $g'(N) > 0$). For example, a doctor in the first quartile has a smaller opportunity to learn from patients than a doctor in the second quartile because, by construction, has fewer observational cases of bad cholesterol; similarly, a doctor in the second quartile has a smaller opportunity to learn from patients than a doctor in the third quartile.

To check this prediction, in Table 8 we report the average number of tests, over time and by quartile of number of patients suffering from cholesterol. As going from one quartile to the following, the number of tests delivered decreases. Although the differences are generally small, they are larger when the switch is from a lower to an upper dose; in contrast, when the switch is from an upper to a lower dose, not only are the differences smaller, but also in a less clear direction (i.e., some times there is an increase in number of test as quantile of patient increases and some times a decrease). Although our model tells nothing about this direction, we think reasonable that there is no concern when moving to a lower dose, but there is when the intake increases.

When physicians have no way to learn from their patients (recall that $g(0) = 0$), they can only learn from official documentation, i.e. clinical trial. Confirming this feature of our model, the average number of tests is higher before evidence against the belief of statin as the cause of liver demages is provided (i.e., the difference is higher before 2006), to decrease the first time when studies began to suggest that statins might not be the cause of liver function abnormalities in

patients taking statins in 2006, and reduce further when it was finally established that statins may actually help patients with liver function abnormalities.

Our model does not say anything on a threshold for $g(N) >> 0$. We rely on quartiles for simplicity of exposition, which conveys the effects across the distribution adequately. With this quartile based split, the most striking results are for 'extreme physicians', that is, those without many opportunities to learn from their patients (i.e., $g(0) = 0$) and those who can learn much (i.e., $g(N) = 1$); in between, the differences are more nuanced.

Interestingly, the differences (across different quantiles) in differences (across eras) suggest that the learning process is decreasing in quantiles of observed patients, thus physicians at first quartile learned more after 2006 than physicians at second quartile. Some exceptions are observed only when the third quartile is involved. Although this is not strictly required by our model, we view this result as reinforcing the evidence in favour of the mechanism we proposed regarding the physicians' learning.

6 Conclusion

In this paper we investigate the relation between technological advance, in the form of an information shock, and changes in physicians and patient behavior. We focus on the case of statins, a medication used to manage high cholesterol levels, where newly produced evidence has altered the scientific consensus regarding side effect profiles.

Early common wisdom held that statins may cause liver damage and therefore patients should be regularly tested for changes in liver enzyme levels. In 2010 GREACE, a large and influential randomized controlled trial, suggested instead that the medication could be continued despite elevated liver enzymes, and that patients need not be tested regularly for liver damage. We exploit this

major informational technological shock to test how physicians prescription and testing behavior and patients adherence to therapy changed in response.

We test our model using a unique dataset representative of the Italian population, that links patients to doctors over the period 2003-2014. We account for the possible non-random sorting of patients into treatment by exploiting an instrument which is assigned effectively at random. We employ a non-parametric bounding approach that takes into account the selection mechanism (unlike ordinary least squares) and permits conclusions about the entire population (unlike standard instrumental variables estimation). Our results show that doctors responded promptly to this technological shock.

References

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables (Disc: p456-472). *Journal of the American Statistical Association 91*, 444–455.

Atella, V., F. Belotti, and D. Depalo (2017, September). Drug therapy adherence and health outcomes in the presence of physician and patient unobserved heterogeneity. *Health Economics 26*, 106–126.

Athyros, V. G., K. Tziomalos, T. D. Gossios, T. Griva, P. Anagnostis, K. Kargiotis, E. D. Pagourelias, E. Theocharidou, A. Karagiannis, and D. P. Mikhailidis (2010, dec). Safety and efficacy of long-term statin treatment for cardiovascular events in patients with coronary heart disease and abnormal liver tests in the Greek Atorvastatin and Coronary Heart Disease Evaluation (GREACE) Study: a post-hoc analysis. *The Lancet 376*(9756), 1916–1922.

Bays, H. (2006). Statin Safety: An Overview and Assessment of the Data-2005. *American Journal of Cardiology 97*(8 SUPPL. 1).

Bhardwaj, S. S. and N. Chalasani (2007, aug). Lipid-Lowering Agents That Cause Drug-Induced Hepatotoxicity. *Clinics in Liver Disease 11*(3), 597–613.

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2008, May). Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization. *American Economic Review 98*(2), 351–56.

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2012). Treatment effect bounds: An application to SwanGanz catheterization. *Journal of Econometrics 168*(2), 223–243.

Blundell, R., L. Dearden, and B. Sianesi (2005). Evaluating the effect of education on earnings: models, methods and results from the national child development survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 168*(3), 473–512.

Bound, J., D. Jaeger, and R. Baker (1995). Problems with instrumental variable estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*, 443–450.

Bradford, R. H., C. L. Shear, A. N. Chremos, C. Dujovne, M. Downton, F. A. Franklin, A. L. Gould, M. Hesney, J. Higgins, and D. P. Hurley (1991, jan). Expanded Clinical Evaluation of Lovastatin (EXCEL) study results. I. Efficacy in modifying plasma lipoproteins and adverse event profile in 8245 patients with moderate hypercholesterolemia. *Archives of internal medicine 151*(1), 43–9.

Calderon, R. M., L. X. Cubeddu, R. B. Goldberg, and E. R. Schiff (2010, apr). Statins in the treatment of dyslipidemia in the presence of elevated liver aminotransferase levels: a therapeutic dilemma. *Mayo Clinic proceedings 85*(4), 349–56.

Charlson, M., P. Pompei, K. Ales, and C. R. MacKenzie (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis 40*(5), 373–383.

Chen, X., C. A. Flores, and A. Flores-Lagunes (2017). Going beyond LATE: Bounding Average Treatment Effects of Job Corps Training. *Journal of Human Resources*, n/a–n/a.

Chernozhukov, V. (2015). Implementing intersection bounds in stata. *Stata Journal 15*(1), 21–44(24).

Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica 81*(2), 667–737.

Clarke, A. and P. Mills (2006, oct). Atorvastatin associated liver disease. *Digestive and Liver Disease 38*(10), 772–777.

Cohen, D. E., F. A. Anania, N. Chalasani, and National Lipid Association Statin Safety Task Force Liver Expert Panel (2006, apr). An Assessment of Statin Safety by Hepatologists. *The American Journal of Cardiology 97*(8), S77–S81.

Conforti, A., L. Magro, U. Moretti, S. Scotto, D. Motola, F. Salvo, B. Ros, and R. Leone (2006). Fluvastatin and hepatic reactions: a signal from spontaneous reporting in Italy. *Drug safety 29*(12), 1163–72.

Cramer, J., A. Roy, A. Burrell, and ... (2008). Medication compliance and persistence: Terminology and definitions. *Value in Health 11*.

Deaton, A. and N. Cartwright (2017). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*.

Depalo, D. (2017). Identification Issues in the Public/Private Wage Gap with an Application to Italy. *Journal of Applied Econometrics*, na–na.

Fabiani, L., M. Scatigna, K. Panopoulou, A. Sabatini, E. Sessa, F. Donato, M. Marchi, R. Nardi, C. Niccolai, F. Samani, and G. Ventriglia (2004). Health search: istituto di ricerca della societá italiana di medicina generale: la realizzazione di un database per la ricerca in medicina generale. *Epidemiol and Prev 28*, 156–162.

Frolkis, J. P., G. L. Pearce, V. Nambi, S. Minor, and D. L. Sprecher (2002, dec). Statins do not meet expectations for lowering low-density lipoprotein

cholesterol levels when used in clinical practice. *The American Journal of Medicine 113*(8), 625–629.

Gillett, R. C. and A. Norrell (2011, mar). Considerations for safe use of statins: liver enzyme abnormalities and muscle toxicity. *American family physician 83*(6), 711–6.

Heckman, J. (1997). Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *The Journal of Human Resources 32*(3), 441–462.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006, August). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

Hughes, D., A. Bagust, A. Haycox, and T. Walley (2001). The impact of non-compliance on the cost-effectiveness of pharmaceuticals: a review of the literature. *Health Economics 10*, 601–615.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–75.

Imbens, G. W. and C. F. Manski (2004, November). Confidence Intervals for Partially Identified Parameters. *Econometrica 72*(6), 1845–1857.

Imbens, G. W. and D. B. Rubin (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies 64*(4), 555–574.

Khorashadi, S., N. K. Hasson, and R. C. Cheung (2006, jul). Incidence of Statin Hepatotoxicity in Patients With Hepatitis C. *Clinical Gastroenterology and Hepatology 4*(7), 902–907.

Kreider, B. and J. V. Pepper (2007, June). Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors. *Journal of the American Statistical Association 102*, 432–441.

Lewis, J. H., M. E. Mortensen, S. Zweig, M. J. Fusco, J. R. Medoff, R. Belder, and Pravastatin in Chronic Liver Disease Study Investigators (2007, nov). Efficacy and safety of high-dose pravastatin in hypercholesterolemic patients with well-compensated chronic liver disease: Results of a prospective, randomized, double-blind, placebo-controlled, multicenter trial. *Hepatology 46*(5), 1453–1463.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association 90*(431), 1112–1121.

Manski, C. (2009). *Identification for Prediction and Decision*. Harvard University Press.

Manski, C. F. (1990, May). Nonparametric Bounds on Treatment Effects. *American Economic Review 80*(2), 319–23.

Manski, C. F. (2013, December). *Public Policy in an Uncertain World: Analysis and Decisions*. Number 9780674066892 in Economics Books. Harvard University Press.

Manski, C. F. and J. V. Pepper (2000, July). Monotone Instrumental Variables, with an Application to the Returns to Schooling. *Econometrica 68*(4), 997–1012.

Maron, D., S. Fazio, and M. Linton (2000). Current perspectives on statins. *Circulation 101*, 207–213.

McCarthy, I., D. L. Millimet, and M. Roy (2015, June). Bounding treatment effects: A command for the partial identification of the average treatment

effect with endogenous and misreported treatment assignment. *Stata Journal 15*(2), 411–436.

McKenney, J. M., M. H. Davidson, T. A. Jacobson, J. R. Guyton, and National Lipid Association Statin Safety Assessment Task Force (2006, apr). Final Conclusions and Recommendations of the National Lipid Association Statin Safety Assessment Task Force. *The American Journal of Cardiology 97*(8), S89–S94.

Mölgaard, J., B. L. Lundh, H. von Schenck, and A. G. Olsson (1991, dec). Long-term efficacy and safety of simvastatin alone and in combination therapy in treatment of hypercholesterolaemia. *Atherosclerosis 91 Suppl*, S21–8.

Mourifie, I. and Y. Wan (2016). Testing local average treatment effect assumptions. *Review of Economics and Statistics na*, n/a–n/a.

Oreopoulos, P. (2006, March). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review 96*(1), 152–175.

Pastori, D., L. Polimeni, F. Baratta, A. Pani, M. Del Ben, and F. Angelico (2015, jan). The efficacy and safety of statins for the treatment of non-alcoholic fatty liver disease. *Digestive and Liver Disease 47*(1), 4–11.

Pepper, J. V. (2000). The intergenerational transmission of welfare receipt: A nonparametric bounds analysis. *Review of Economics and Statistics 82*(3), 472–488.

Plakogiannis, R. and H. Cohen (2007, jan). Optimal low-density lipoprotein cholesterol lowering–morning versus evening statin administration. *The Annals of pharmacotherapy 41*(1), 106–10.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Rzouq, F. S., M. L. Volk, H. H. Hatoum, S. K. Talluri, R. R. Mummadi, and G. K. Sood (2010, aug). Hepatotoxicity Fears Contribute to Underutilization of Statin Medications by Primary Care Physicians. *The American Journal of the Medical Sciences 340*(2), 89–93.

Scandinavian Simvastatin Survival Study Group (1994, nov). Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *The Lancet 344*(8934), 1383–1389.

Segarra-Newnham, M., D. Parra, and E. M. Martin-Cooper (2007, jun). Effectiveness and Hepatotoxicity of Statins in Men Seropositive for Hepatitis C Virus. *Pharmacotherapy 27*(6), 845–851.

Shaikh, A. and E. Vytlacil (2011). Partial identification in triangular systems of equations with binary dependent variables. *Econometrica 79*(3), pp. 949–955.

Sniderman, A. D. (2004, nov). Is there value in liver function test and creatine phosphokinase monitoring with statin use? *The American Journal of Cardiology 94*(9), 30–34.

Vytlacil, E. (2002, jan). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica 70*(1), 331–341.

Table 1: Statin conversion table

Class	% LDL red.	Simva	Ator	Rosu
1	<24	5	-	-
2	25–32	10	-	-
3	31–39	20	10	-
4	37–45	40	20	5
5	48–52	80	40	10
6	55–60	-	80	20
7	60–63	-	-	40

Table 2: Descriptive statistics

	Women			Men		
Variabile	Mean	S.D.	Median	Mean	S.D.	Median
Δ Chol.	-0.03812	0.19161	-0.01911	-0.03776	0.18461	-0.02010
LDL	120.70131	36.26905	115.77856	110.84665	34.90859	106.31404
Adherence	0.64299	0.28583	0.62921	0.69568	0.28073	0.67416
ALP	0.06080	0.25567	0.00000	0.05228	0.24262	0.00000
ALT	0.29048	0.50879	0.00000	0.28997	0.51597	0.00000
AST	0.27201	0.49636	0.00000	0.26991	0.50155	0.00000
Albumin	0.02888	0.17698	0.00000	0.02967	0.18138	0.00000
Birilubin	0.02181	0.15431	0.00000	0.02190	0.15578	0.00000
Birilubin (Tot & Fract.)	0.04222	0.21578	0.00000	0.04308	0.22302	0.00000
Charlson ind.	1.00882	1.15930	1.00000	1.22023	1.26860	1.00000
Equip.	4.49154	0.64965	4.00000	4.61510	0.69400	5.00000
Diabetes	0.24669	0.43109	0.00000	0.29706	0.45696	0.00000
Hypert.	0.72070	0.44865	1.00000	0.68348	0.46512	1.00000
Congestive heart failure	0.02697	0.16200	0.00000	0.03852	0.19246	0.00000
Atrial fibrillation	0.05063	0.21925	0.00000	0.06261	0.24225	0.00000
Vasc. Deas.	0.01009	0.09994	0.00000	0.01934	0.13772	0.00000
PCI	0.00132	0.03629	0.00000	0.00511	0.07132	0.00000
Ischemic heart	0.01392	0.11714	0.00000	0.01733	0.13050	0.00000
Other Heart	0.13643	0.34324	0.00000	0.31022	0.46258	0.00000
Age	68.94675	9.31172	70.00000	65.53925	10.28141	66.00000
North W	0.24379	0.42937	0.00000	0.26874	0.44331	0.00000
Nort E	0.21095	0.40799	0.00000	0.21464	0.41057	0.00000
Center	0.18273	0.38645	0.00000	0.17311	0.37834	0.00000
South	0.24468	0.42990	0.00000	0.23888	0.42640	0.00000
simvastatin	0.44051	0.49645	0.00000	0.39973	0.48984	0.00000
atorvastatin	0.34151	0.47422	0.00000	0.38324	0.48618	0.00000
rosuvastatin	0.21797	0.41287	0.00000	0.21703	0.41223	0.00000

Table 3: Number of prescription by active ingredient and dosage

		Statin		
Variable	**mg**	**Simva.**	**Ator.**	**Rosu.**
Men	5	0	0	18705
	10	14282	105253	132127
	20	246066	166173	26579
	40	70926	42047	2455
	80	0	4131	0
Women	5	0	0	25440
	10	24684	137788	158079
	20	331071	164907	24892
	40	68835	24946	1684
	80	0	1525	0

Table 4: Reduction of Cholesterol and Adherence

	BSV		Point		Averages		
Statin	Lower	Upper	OLS	IV	$[Y\|D=0]$	$[Y\|D=1]$	ΔY
All							
Simva.	-0.43697	-0.01120	-0.01251	-0.01308	-0.04134	-0.03007	-0.01127
Ator.	-0.46370	-0.01962	-0.02165	-0.01657	-0.04707	-0.02688	-0.02019
Rosu.	0.00123	-0.02041	-0.02530	0.00980	-0.05739	-0.03277	-0.02462
All	-0.47358	-0.01837	-0.02006	-0.01417	-0.04830	-0.02930	-0.01899
Before 2006							
Simva.	-0.46090	-0.01924	-0.01806	-0.02855	-0.04528	-0.02740	-0.01788
Ator.	-0.51883	-0.02504	-0.02597	-0.02141	-0.06181	-0.03687	-0.02493
Rosu.	-0.63957	-0.06442	-0.05002	-0.14422	-0.14016	-0.08948	-0.05068
All	-0.51038	-0.03818	-0.03431	-0.05770	-0.06857	-0.03488	-0.03368
2006-2011							
Simva.	-0.62380	-0.00939	-0.01173	-0.00578	-0.03910	-0.02887	-0.01023
Ator.	-0.46324	-0.01979	-0.01995	-0.02327	-0.03948	-0.02035	-0.01914
Rosu.	-0.66817	-0.02599	-0.02913	-0.00823	-0.06408	-0.03580	-0.02828
All	-0.49773	-0.02137	-0.02228	-0.02125	-0.04809	-0.02694	-0.02114
After 2011							
Simva.	-0.44628	-0.00992	-0.01190	-0.01282	-0.04269	-0.03242	-0.01027
Ator.	-0.49134	-0.02170	-0.02297	-0.03021	-0.05001	-0.02915	-0.02086
Rosu.	-0.71471	-0.01835	-0.02091	0.00281	-0.04214	-0.02118	-0.02095
All	-0.48154	-0.01693	-0.01670	-0.02839	-0.04500	-0.02944	-0.01556

Table 5: Reduction of Cholesterol and change in therapy (more equivalence)

	BSV		Point		Averages		
Statin	Lower	Upper	OLS	IV	$[Y\|D=0]$	$[Y\|D=1]$	ΔY
				All			
Simva.	-0.49922	-0.00236	0.00190	-0.02847	-0.03266	-0.03525	0.00260
Ator.	-0.68968	-0.00259	0.00748	-0.00218	-0.03090	-0.03850	0.00759
Rosu.	0.00139	0.00594	0.00470	0.02398	-0.04483	-0.04983	0.00500
All	-0.48325	-0.00102	0.00321	-0.00874	-0.03634	-0.03984	0.00349
				Before 2006			
Simva.	-0.56281	-0.00646	-0.00489	-0.05249	-0.03603	-0.03081	-0.00521
Ator.	-0.69352	-0.00162	0.00016	-0.02591	-0.04292	-0.04351	0.00059
Rosu.	0.00768	0.01504	0.05387	0.56251	-0.06833	-0.11806	0.04973
All	-0.55212	-0.00959	-0.00505	-0.06541	-0.04779	-0.04299	-0.00480
				2006-2011			
Simva.	-0.53375	-0.01454	-0.00753	-0.10426	-0.04007	-0.03256	-0.00752
Ator.	-0.52547	-0.00307	0.00031	-0.06596	-0.02796	-0.02819	0.00023
Rosu.	-0.67721	-0.00358	0.00355	-0.15750	-0.04927	-0.05424	0.00498
All	-0.53566	-0.00835	-0.00510	-0.05547	-0.04080	-0.03599	-0.00481
				After 2011			
Simva.	-0.53699	-0.00877	-0.00644	-0.06850	-0.04312	-0.03634	-0.00678
Ator.	-0.58536	-0.00187	-0.00053	-0.04750	-0.04009	-0.03917	-0.00092
Rosu.	-0.63182	-0.00181	-0.00212	-0.25837	-0.03649	-0.03404	-0.00245
All	-0.54539	-0.00980	-0.00833	-0.03535	-0.04589	-0.03720	-0.00869

Table 6: Characterization of subpopulations

	Never				D=0, Compliers				D=1,Compliers				Always			
Variable	All	Pre	Dur.	Post	All	Pre	Dur.	Post	All	Pre	Dur.	Post	All	Pre	Dur.	
age	67.42	64.78	67.04	68.86	68.02	64.93	67.09	69.01	67.06	64.03	66.15	67.72	66.93	63.89	65.95	6
albumina num	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.04	0.03	0.03	0.04	0.03	
alp num	0.06	0.06	0.06	0.06	0.05	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.06	0.06	
alt gpt num	0.28	0.30	0.30	0.27	0.27	0.29	0.28	0.26	0.31	0.33	0.34	0.32	0.30	0.34	0.32	
anem	0.08	0.03	0.06	0.11	0.07	0.03	0.06	0.10	0.09	0.03	0.07	0.12	0.08	0.04	0.07	
asma	0.06	0.04	0.05	0.07	0.05	0.03	0.05	0.07	0.06	0.04	0.05	0.09	0.06	0.04	0.05	
ast got num	0.27	0.29	0.29	0.25	0.24	0.27	0.26	0.24	0.30	0.32	0.33	0.29	0.27	0.33	0.30	
bilirub tot e fraz num	0.05	0.03	0.04	0.05	0.04	0.02	0.04	0.04	0.05	0.02	0.04	0.05	0.04	0.03	0.04	
bilirub tot num	0.02	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.03	0.02	
bmi	28.58	28.36	28.62	28.50	28.62	28.60	28.73	28.57	28.58	28.23	28.69	28.66	28.62	28.54	28.61	2
bpco	0.07	0.05	0.07	0.08	0.07	0.05	0.07	0.08	0.08	0.06	0.06	0.09	0.07	0.06	0.07	
col hdl	1.41	1.43	1.41	1.42	1.44	1.44	1.41	1.44	1.40	1.44	1.39	1.40	1.40	1.43	1.38	
col ldl	2.98	3.29	3.04	2.84	2.97	3.33	3.06	2.85	3.09	3.54	3.22	3.12	3.10	3.54	3.24	
col tot	5.12	5.53	5.19	4.95	5.13	5.57	5.22	4.99	5.29	5.85	5.42	5.26	5.30	5.85	5.46	
compliance	0.64	0.54	0.66	0.69	0.63	0.53	0.66	0.68	0.70	0.62	0.74	0.74	0.70	0.61	0.74	
du area1	0.22	0.26	0.26	0.23	0.30	0.25	0.26	0.29	0.22	0.27	0.26	0.21	0.29	0.25	0.26	
du area2	0.16	0.20	0.20	0.22	0.28	0.26	0.24	0.20	0.15	0.20	0.19	0.22	0.27	0.22	0.23	
du area3	0.21	0.20	0.16	0.17	0.15	0.14	0.19	0.19	0.21	0.20	0.19	0.16	0.14	0.17	0.17	
du area4	0.26	0.22	0.25	0.25	0.22	0.26	0.23	0.23	0.26	0.20	0.24	0.25	0.24	0.27	0.24	
du area5	0.15	0.12	0.12	0.13	0.06	0.09	0.09	0.08	0.16	0.12	0.12	0.16	0.07	0.09	0.10	
du bypass pci	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.01	
du decesso	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
du diab	0.26	0.22	0.28	0.28	0.25	0.22	0.27	0.27	0.29	0.21	0.30	0.29	0.27	0.22	0.29	
du fib atr	0.05	0.03	0.05	0.07	0.06	0.03	0.05	0.07	0.05	0.03	0.04	0.07	0.06	0.04	0.05	
du hyper	0.71	0.62	0.71	0.74	0.71	0.63	0.70	0.73	0.71	0.61	0.70	0.74	0.69	0.60	0.69	
du ictus	0.12	0.08	0.12	0.16	0.13	0.09	0.12	0.14	0.14	0.09	0.14	0.19	0.15	0.09	0.14	
du isch	0.01	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.02	0.01	0.02	
du malcor	0.18	0.22	0.21	0.19	0.17	0.23	0.21	0.19	0.25	0.25	0.28	0.24	0.26	0.27	0.31	
du sc card	0.03	0.02	0.03	0.04	0.03	0.02	0.03	0.03	0.04	0.03	0.03	0.04	0.04	0.03	0.03	
du vasc d	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.02	0.01	0.02	0.02	0.02	
ducharlson1	0.40	0.49	0.39	0.34	0.40	0.48	0.40	0.37	0.37	0.49	0.36	0.31	0.36	0.47	0.36	
ducharlson2	0.32	0.30	0.33	0.32	0.32	0.30	0.32	0.32	0.32	0.29	0.33	0.32	0.32	0.31	0.33	
ducharlson3	0.17	0.13	0.17	0.19	0.17	0.14	0.17	0.18	0.18	0.14	0.19	0.20	0.18	0.14	0.18	
ducharlson4	0.11	0.08	0.11	0.15	0.11	0.07	0.11	0.13	0.13	0.08	0.12	0.17	0.14	0.08	0.13	
female	0.54	0.55	0.53	0.54	0.55	0.54	0.53	0.55	0.53	0.55	0.52	0.54	0.53	0.53	0.50	
fumo	0.03	0.05	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.04	0.03	
fumo ex	0.04	0.03	0.03	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.04	0.04	0.03	0.03	
fumo no	0.05	0.05	0.04	0.06	0.05	0.05	0.04	0.05	0.05	0.04	0.04	0.06	0.04	0.05	0.03	
i bmi	28.33	27.88	28.28	28.39	28.34	28.06	28.37	28.46	28.30	27.96	28.31	28.53	28.31	27.95	28.24	2
i p max	135.53	137.80	136.50	134.76	136.27	138.46	136.23	134.85	135.78	137.53	136.64	134.27	135.89	137.38	135.94	1
i p min	79.42	80.95	79.90	78.41	79.36	81.13	79.67	78.42	79.51	81.02	79.93	78.52	79.34	80.87	79.52	7
mrge	0.19	0.10	0.16	0.26	0.17	0.09	0.16	0.22	0.21	0.11	0.18	0.29	0.19	0.11	0.17	
obes	0.08	0.06	0.08	0.10	0.10	0.06	0.08	0.10	0.08	0.06	0.09	0.11	0.09	0.06	0.09	
parkinson	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	
q a	9.51	8.56	9.74	10.76	10.02	8.52	9.90	10.04	10.34	8.78	10.54	12.37	10.83	9.52	10.86	1
q a 2	0.34	0.36	0.39	0.30	0.31	0.35	0.35	0.29	0.38	0.38	0.42	0.35	0.35	0.39	0.40	
s a	64.21	54.28	67.50	69.49	62.55	52.42	63.40	66.42	70.39	55.70	73.98	78.94	69.78	58.26	74.18	7
s a 2	6.91	7.40	8.14	6.08	6.43	7.20	7.40	5.93	7.54	7.67	8.68	6.86	7.43	7.91	8.48	
s t	218.16	248.75	228.68	215.73	209.19	239.46	226.66	213.45	242.92	254.08	255.51	231.79	247.17	264.78	263.84	2
s tot	282.37	303.03	296.18	285.22	271.75	291.88	290.06	279.88	313.31	309.78	329.49	310.73	316.96	323.04	338.02	3
sport leggero	0.07	0.08	0.07	0.08	0.08	0.09	0.08	0.07	0.07	0.08	0.07	0.08	0.08	0.08	0.09	
sport medio	0.01	0.02	0.01	0.01	0.02	0.03	0.02	0.01	0.01	0.03	0.01	0.01	0.02	0.02	0.02	
sport no	0.08	0.10	0.09	0.09	0.09	0.08	0.08	0.08	0.09	0.08	0.09	0.10	0.08	0.08	0.07	
sport pesante	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
tiroide	0.21	0.12	0.18	0.25	0.19	0.12	0.18	0.23	0.21	0.13	0.17	0.27	0.19	0.12	0.18	
trigli	1.65	1.84	1.70	1.58	1.61	1.85	1.71	1.58	1.77	2.01	1.82	1.69	1.78	2.09	1.89	
tum	0.11	0.08	0.11	0.14	0.12	0.09	0.11	0.13	0.11	0.07	0.10	0.13	0.11	0.07	0.10	

Table 7: Pattern of behaviour of liver test

Number	Overall	Lower dose	Higher dose
	All		
1	0.07212 ***	0.00930 ***	0.06789 ***
2	0.08176 ***	0.01049 ***	0.07588 ***
3	0.07025 ***	0.00807 ***	0.06623 ***
4	0.12224 ***	0.01636 ***	0.11572 ***
5	0.12011 ***	0.01623 ***	0.11528 ***
	Before 2006		
1	0.07886 ***	0.00849 ***	0.07493 ***
2	0.08679 ***	0.01040 ***	0.07847 ***
3	0.08698 ***	0.01106 ***	0.07847 ***
4	0.14324 ***	0.02138 ***	0.13541 ***
5	0.16685 ***	0.02690 ***	0.16181 ***
	2006-2010		
1	0.08341 ***	0.00849 ***	0.08015 ***
2	0.08738 ***	0.01081 ***	0.08301 ***
3	0.08126 ***	0.00956 ***	0.07738 ***
4	0.13702 ***	0.01839 ***	0.13055 ***
5	0.13595 ***	0.01750 ***	0.12986 ***
	After 2010		
1	0.06654 ***	0.01071 ***	0.06123 ***
2	0.08120 ***	0.00991 ***	0.07674 ***
3	0.06299 ***	0.00618 ***	0.06022 ***
4	0.10443 ***	0.01296 ***	0.09897 ***
5	0.09365 ***	0.01250 ***	0.08811 ***

Table 8: Pattern of behaviuour of liver test - by quartile

# Test	All			Quartile											
				1			2			3			4		
	Overall	Lower	Higher	Overall	Lower	Higher	Overall	Lower	Higher	Overall	Lower	Higher	Overall	Lower	Higher
							All								
1	0.072***	0.009***	0.068***	0.099***	0.009***	0.096***	0.070***	0.010***	0.065***	0.069***	0.008***	0.065***	0.067***	0.010***	0.062***
2	0.082***	0.010***	0.076***	0.076***	0.008***	0.072***	0.083***	0.010***	0.078***	0.078***	0.010***	0.074***	0.086***	0.012***	0.080***
3	0.070***	0.008***	0.066***	0.053***	0.009***	0.049***	0.068***	0.008***	0.065***	0.067***	0.007***	0.064***	0.079***	0.008***	0.075***
4	0.122***	0.016***	0.116***	0.109***	0.013***	0.104***	0.125***	0.016***	0.119***	0.127***	0.018***	0.119***	0.123***	0.016***	0.116***
5	0.120***	0.016***	0.115***	0.111***	0.016***	0.105***	0.115***	0.017***	0.108***	0.136***	0.016***	0.130***	0.117***	0.016***	0.110***
							Before 2006								
1	0.079***	0.008***	0.075***	0.117***	0.018***	0.112***	0.062***	0.005	0.061***	0.093***	0.005	0.092***	0.059***	0.010**	0.055***
2	0.087***	0.010***	0.078***	0.091***	0.010***	0.088***	0.085***	0.007***	0.083***	0.081***	0.009***	0.078***	0.092***	0.013***	0.087***
3	0.087***	0.011***	0.078***	0.056***	0.013***	0.050***	0.080***	0.016***	0.074***	0.075***	0.003	0.075***	0.111***	0.013***	0.107***
4	0.143***	0.021***	0.135***	0.139***	0.021***	0.134***	0.138***	0.017***	0.133***	0.164***	0.029***	0.156***	0.136***	0.019***	0.130***
5	0.167***	0.027***	0.162***	0.153***	0.029***	0.144***	0.149***	0.023***	0.143***	0.180***	0.035***	0.171***	0.169***	0.022***	0.163***
							2006-2010								
1	0.083***	0.008***	0.080***	0.116***	0.010***	0.113***	0.085***	0.009***	0.082***	0.083***	0.009***	0.079***	0.072***	0.007***	0.069***
2	0.087***	0.011***	0.083***	0.082***	0.006***	0.080***	0.089***	0.012***	0.084***	0.081***	0.010***	0.077***	0.093***	0.012***	0.089***
3	0.081***	0.010***	0.077***	0.071***	0.011***	0.066***	0.079***	0.008***	0.076***	0.082***	0.010***	0.078***	0.086***	0.009***	0.083***
4	0.137***	0.018***	0.131***	0.129***	0.017***	0.123***	0.139***	0.020***	0.132***	0.136***	0.021***	0.129***	0.140***	0.017***	0.134***
5	0.136***	0.017***	0.130***	0.117***	0.019***	0.110***	0.146***	0.023***	0.138***	0.145***	0.014***	0.141***	0.133***	0.017***	0.127***
							After 2010								
1	0.067***	0.011***	0.061***	0.084***	0.008***	0.081***	0.062***	0.013***	0.056***	0.058***	0.007***	0.055***	0.069***	0.013***	0.062***
2	0.081***	0.010***	0.077***	0.068***	0.007***	0.065***	0.080***	0.009***	0.076***	0.081***	0.010***	0.077***	0.086***	0.011***	0.081***
3	0.063***	0.006***	0.060***	0.053***	0.006***	0.050***	0.062***	0.007***	0.059***	0.055***	0.006***	0.052***	0.072***	0.006***	0.069***
4	0.104***	0.013***	0.099***	0.084***	0.005**	0.082***	0.108***	0.014***	0.102***	0.107***	0.013***	0.102***	0.108***	0.015***	0.102***
5	0.094***	0.012***	0.088***	0.092***	0.009***	0.089***	0.087***	0.012***	0.082***	0.112***	0.014***	0.106***	0.089***	0.013***	0.083***

A Testing Monotonicity and Exogeneity

We also tested the non-rejection of the validity of exogeneity and monotonicity assumptions using a test proposed by Mourifie and Wan (2016).

Notice that $E[Y|D=0,Z=0] = \frac{\pi_c}{\pi_c+\pi_{nt}} Y_c + \frac{\pi_{nt}}{\pi_c+\pi_{nt}} Y_{nt}$. This implies the following testable implications when $LATE_c$ is positive: $E[Y|D=1,Z=0] \geq E[Y|D=0,Z=0]$ (related to the always takers part) and $E[Y|D=1,Z=1] \geq E[Y|D=0,Z=1]$ (related to the never takers part). Focus on the always takers. In this case, fixing $Z=0$, a positive treatment effect makes $E[Y_1] \geq E[Y_0]$. As a consequence, if $LATE_C \geq 0$, rejecting this hypothesis is strong evidence against them (Chen et al., 2017).

When the treatment effect $LATE_C \leq 0$, then $E[Y_0|at] \geq E[Y_1|at]$. Therefore, non-rejecting the null hypothesis that $E[Y|D=1,Z=0] \geq E[Y|D=0,Z=0]$ when $LATE_c \leq 0$ is a strong supporting argument in favour of the working hypotheses, because it implies that $E[Y_0|at] \geq E[Y_1|at] = E[Y|D=1,Z=0] \geq E[Y|D=0,Z=0]$.

In our exercise, for the case of lower equipotency we obtain that $E[Y|D=1,Z=0] = -.03650743$,$E[Y|D=0,Z=0] = -.03840154$, and $E[Y|D=1,Z=1] = -.0364091$ and $E[Y|D=0,Z=1] = -.04198889$; for the case of higher equipotency, the $E[Y|D=1,Z=0] = -.03575313$,$E[Y|D=0,Z=0] = -.03908531$, and $E[Y|D=1,Z=1] = -.03685148$ and $E[Y|D=0,Z=1] = -.04085401$, therefore in both cases $E[Y|D=1,Z=0] \geq E[Y|D=0,Z=0]$, and $E[Y|D=1,Z=1] \geq E[Y|D=0,Z=1]$ as required by assumption 3 of the bounds in Chen et al. (2017).

The economic content of this condition is non trivial: it suggests that if the always takers did not take pills, their reduction of cholesterol in the population would be the lowest. As for the never takers, they can substitute pills with other behviour, like physical activity (Atella et al., 2017).

B Inference

Bounds in eq. 23 involve minima and maxima. In finite sample these estimators are biased, typically providing conservative (i.e., smaller) bounds. Corrections have been proposed by Chernozhukov et al. (2013); Kreider and Pepper (2007). To implement the method in Kreider and Pepper (2007), define T_n the sample analog of the consistent estimator of the parameter θ, by definition the bias $b_n = E[T_n] - \theta$. The nonparametric bootstrap delivers $\hat{b} = E^*[T_n] - T_n$, where $E^*[\cdot]$ is the expectations with respect to the bootstrap distribution. The bootstrap bias-corrected estimator can now be calculated as $T_n^c = T_n - \hat{b} = 2T_n - E^*[T_n]$. Kreider and Pepper (2007) provide evidence in favour of this approach. The method by Chernozhukov et al. (2013) is based on precision-corrected estimate of the sample analog estimator $\theta\hat{(p)} + k(p)\, s(x)$, with $s(x)$ the standard error of $\theta\hat{(x)}$, and $k(p)$ is a critical value that is based on an adaptive inequality selection procedure proposed on purpose. In our application we rely on Kreider and Pepper (2007) for several reasons: 1) it is easier to understand and to implement (notice however that Chernozhukov, 2015 provide a Stata-routine for the method); 2) it is computationally much faster than the alternative Chernozhukov et al., 2013; Chernozhukov, 2015; 3) the correction in Kreider and Pepper (2007) performs well based on Montecarlo evidence provided by the authors. 4) with our sample size, the difference between the two methods should be negligible.

Based on the bootstrap bias-corrected estimator, Kreider and Pepper (2007) provide the associated confidence interval. McCarthy et al. (2015) propose to use either Kreider and Pepper (2007) altogether or the method of Kreider and Pepper (2007) for the finite-sample correction and the asymptotic inference suggested by Imbens and Manski (2004).